Density estimation using Dirichlet kernels

Frédéric Ouimet[a,1,*]

[a] *California Institute of Technology, Pasadena, USA.*

Abstract

In this paper, we introduce Dirichlet kernels for the estimation of multivariate densities supported on the d-dimensional simplex. These kernels generalize the beta kernels from Brown & Chen (1999); Chen (1999, 2000a); Bouezmarni & Rolin (2003), originally studied in the context of smoothing for regression curves. We prove various asymptotic properties for the estimator : bias, variance, mean squared error, mean integrated squared error, asymptotic normality and uniform strong consistency. In particular, the asymptotic normality and uniform strong consistency results are completely new, even for the case $d = 1$ (beta kernels). These new kernel smoothers can be used for density estimation of compositional data. The estimator is simple to use, free of boundary bias, allocates non-negative weights everywhere on the simplex, and achieves the optimal convergence rate of $n^{-4/(d+4)}$ for the mean integrated squared error.

Keywords: Dirichlet kernels, beta kernels, simplex, density estimation, mean squared error, mean integrated squared error, asymptotic normality, uniform strong consistency
2010 MSC: Primary : 62G07 Secondary : 62G05, 62G20

1. Dirichlet kernel estimators

The d-dimensional simplex and its interior are defined by

$$\mathcal{S} := \big\{\boldsymbol{s} \in [0,1]^d : \|\boldsymbol{s}\|_1 \leq 1\big\} \quad \text{and} \quad \mathrm{Int}(\mathcal{S}) := \big\{\boldsymbol{s} \in (0,1)^d : \|\boldsymbol{s}\|_1 < 1\big\}, \tag{1.1}$$

where $\|\boldsymbol{s}\|_1 := \sum_{i=1}^d |s_i|$. For $\alpha_1, \dots, \alpha_d, \beta > 0$, the density of the Dirichlet$(\boldsymbol{\alpha}, \beta)$ distribution is

$$K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s}) := \frac{\Gamma(\|\boldsymbol{\alpha}\|_1 + \beta)}{\Gamma(\beta)\prod_{i=1}^d \Gamma(\alpha_i)} \cdot (1 - \|\boldsymbol{s}\|_1)^{\beta-1} \prod_{i=1}^d s_i^{\alpha_i - 1}, \quad \boldsymbol{s} \in \mathcal{S}. \tag{1.2}$$

For a given bandwidth parameter $b > 0$, and a sample of i.i.d. vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$ that are F distributed with a density f supported on $\mathcal{S}$, the Dirichlet kernel estimator is

$$\hat{f}_{n,b}(\boldsymbol{s}) := \frac{1}{n}\sum_{i=1}^n K_{\boldsymbol{s}/b+1,(1-\|\boldsymbol{s}\|_1)/b+1}(\boldsymbol{X}_i). \tag{1.3}$$

Throughout the paper, we use the notation

$$[d] := \{1, 2, \dots, d\}. \tag{1.4}$$

Also, as $n \to \infty$ and/or $b \to 0$, we use the standard asymptotic notation $\mathcal{O}(\cdot)$ and $\mathrm{o}(\cdot)$, which implicitly can depend on the density f and the dimension d, but no other variable unless explicitly written as a subscript.

[*] Corresponding author
Email address: ouimetfr@caltech.edu (Frédéric Ouimet)
[1] F. O. is supported by a postdoctoral fellowship from the NSERC (PDF) and the FRQNT (B3X supplement).

2. Main results

For each result stated in this section, one of the following two assumptions will be used.

Assumptions.

- *The density f is Lipschitz continuous on $\mathcal{S}$.* (2.1)
- *The second order partial derivatives of f are (uniformly) continuous on $\mathcal{S}$.* (2.2)

We denote the expectation of $\hat{f}_{n,b}(\boldsymbol{s})$ by

$$f_b(\boldsymbol{s}) := \mathbb{E}[K_{\boldsymbol{s}/b+1,(1-\|\boldsymbol{s}\|_1)/b+1}(\boldsymbol{X})] = \int_{\mathcal{S}} f(\boldsymbol{x}) K_{\boldsymbol{s}/b+1,(1-\|\boldsymbol{s}\|_1)/b+1}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{2.3}$$

Alternatively, notice that if $\boldsymbol{\xi}_{\boldsymbol{s}} \sim \text{Dirichlet}(\boldsymbol{s}/b+1, (1-\|\boldsymbol{s}\|_1)/b+1)$, then we also have the representation

$$f_b(\boldsymbol{s}) = \mathbb{E}[f(\boldsymbol{\xi}_{\boldsymbol{s}})]. \tag{2.4}$$

Proposition 2.1. *Under assumption (2.2), we have, uniformly for $\boldsymbol{s} \in \mathcal{S}$,*

$$f_b(\boldsymbol{s}) = f(\boldsymbol{s}) + b\, g(\boldsymbol{s}) + \mathrm{o}(b), \tag{2.5}$$

as $b \to 0$, where

$$g(\boldsymbol{s}) := \sum_{i=1}^d (1-(d+1)s_i) \frac{\partial}{\partial s_i} f(\boldsymbol{s}) + \frac{1}{2} \sum_{i,j=1}^d s_i (\mathbb{1}_{\{i=j\}} - s_j) \frac{\partial^2}{\partial s_i \partial s_j} f(\boldsymbol{s}). \tag{2.6}$$

Theorem 2.2 (Bias and variance). *Assuming (2.2), we have, uniformly for $\boldsymbol{s} \in \mathcal{S}$,*

$$\mathbb{B}\mathrm{ias}[\hat{f}_{n,b}(\boldsymbol{s})] = f_b(\boldsymbol{s}) - f(\boldsymbol{s}) = b\, g(\boldsymbol{s}) + \mathrm{o}(b). \tag{2.7}$$

Furthermore, for every subsets $\mathcal{J} \subseteq [d]$, denote

$$\psi(\boldsymbol{s}) := \psi_\emptyset(\boldsymbol{s}) \quad \text{and} \quad \psi_{\mathcal{J}}(\boldsymbol{s}) := \Big[(4\pi)^{d-|\mathcal{J}|} \cdot (1-\|\boldsymbol{s}\|_1) \prod_{i \in [d]\backslash \mathcal{J}} s_i\Big]^{-1/2}. \tag{2.8}$$

Then, for any $\boldsymbol{s} \in \mathrm{Int}(\mathcal{S})$, any $\emptyset \neq \mathcal{J} \subseteq [d]$, and any $\boldsymbol{\kappa} \in (0,\infty)^d$, we have, only assuming (2.1),

$$\mathbb{V}\mathrm{ar}(\hat{f}_{n,b}(\boldsymbol{s})) = \begin{cases} n^{-1} b^{-d/2} (\psi(\boldsymbol{s}) f(\boldsymbol{s}) + \mathcal{O}_{\boldsymbol{s}}(b^{1/2})), & \text{if } s_i/b \to \kappa_i \ \forall i \in [d] \text{ and } (1-\|\boldsymbol{s}\|_1)/b \to \infty, \\ n^{-1} b^{-(d+|\mathcal{J}|)/2} \\ \cdot \Big(\prod_{i \in \mathcal{J}} \frac{\Gamma(2\kappa+1)}{2^{2\kappa_i+1} \Gamma^2(\kappa+1)} \cdot \psi_{\mathcal{J}}(\boldsymbol{s}) f(\boldsymbol{s}) + \mathcal{O}_{\boldsymbol{\kappa},\boldsymbol{s}}(b^{1/2}) \Big), & \text{if } s_i/b \to \kappa_i \ \forall i \in \mathcal{J} \text{ and } s_i/b \to \infty \ \forall i \in [d]\backslash\mathcal{J} \text{ and } (1-\|\boldsymbol{s}\|_1)/b \to \infty. \end{cases} \tag{2.9}$$

Corollary 2.3 (Mean squared error). *Under assumption (2.2), we have, for $\boldsymbol{s} \in \mathrm{Int}(\mathcal{S})$,*

$$\mathrm{MSE}(\hat{f}_{n,b}(\boldsymbol{s})) = n^{-1} b^{-d/2} \psi(\boldsymbol{s}) f(\boldsymbol{s}) + b^2 g^2(\boldsymbol{s}) + \mathcal{O}_{\boldsymbol{s}}(n^{-1} b^{-d/2+1/2}) + \mathrm{o}(b^2). \tag{2.10}$$

In particular, if $f(\boldsymbol{s}) \cdot g(\boldsymbol{s}) \neq 0$, the asymptotically optimal choice of b, with respect to MSE, is

$$b_{\mathrm{opt}} = n^{-2/(d+4)} \left[\frac{d}{4} \cdot \frac{\psi(\boldsymbol{s}) f(\boldsymbol{s})}{g^2(\boldsymbol{s})}\right]^{2/(d+4)}, \quad \text{with} \tag{2.11}$$

$$\mathrm{MSE}[\hat{f}_{n,b_{\mathrm{opt}}}] = n^{-4/(d+4)} \left[\frac{1+\frac{d}{4}}{\big(\frac{d}{4}\big)^{\frac{d}{d+4}}}\right] \frac{\big(\psi(\boldsymbol{s}) f(\boldsymbol{s})\big)^{4/(d+4)}}{\big(g^2(\boldsymbol{s})\big)^{-d/(d+4)}} + \mathrm{o}_{\boldsymbol{s}}(n^{-4/(d+4)}), \tag{2.12}$$

and, more generally, if $n^{2/(d+4)} b \to \lambda$ for some $\lambda > 0$, then

$$\mathrm{MSE}[\hat{f}_{n,b}(\boldsymbol{s})] = n^{-4/(d+4)} \big[\lambda^{-d/2} \psi(\boldsymbol{s}) f(\boldsymbol{s}) + \lambda^2 g^2(\boldsymbol{s})\big] + \mathrm{o}_{\boldsymbol{s}}(n^{-4/(d+4)}). \tag{2.13}$$

Theorem 2.4 (Mean integrated squared error)**.** *Under assumption* (2.2)*, we have*

$$\mathrm{MISE}[\hat{f}_{n,b}] = n^{-1}b^{-d/2}\int_{\mathcal{S}}\psi(\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + b^2\int_{\mathcal{S}} g^2(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + \mathrm{o}(n^{-1}b^{-d/2}) + \mathrm{o}(b^2). \tag{2.14}$$

In particular, if $\int_{\mathcal{S}} g^2(\boldsymbol{s})\mathrm{d}\boldsymbol{s} > 0$*, the asymptotically optimal choice of* b*, with respect to* MISE*, is*

$$b_{\mathrm{opt}} = n^{-2/(d+4)}\left[\frac{d}{4}\cdot\frac{\int_{\mathcal{S}}\psi(\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}{\int_{\mathcal{S}} g^2(\boldsymbol{s})\mathrm{d}\boldsymbol{s}}\right]^{2/(d+4)}, \quad \textit{with} \tag{2.15}$$

$$\mathrm{MISE}[\hat{f}_{n,b_{\mathrm{opt}}}] = n^{-4/(d+4)}\left[\frac{1+\frac{d}{4}}{\left(\frac{d}{4}\right)^{\frac{d}{d+4}}}\right]\frac{\left(\int_{\mathcal{S}}\psi(\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right)^{4/(d+4)}}{\left(\int_{\mathcal{S}} g^2(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right)^{-d/(d+4)}} + \mathrm{o}(n^{-4/(d+4)}), \tag{2.16}$$

and, more generally, if $n^{2/(d+4)}\, b \to \lambda$ *for some* $\lambda > 0$*, then*

$$\mathrm{MISE}[\hat{f}_{n,b}] = n^{-4/(d+4)}\left[\lambda^{-d/2}\int_{\mathcal{S}}\psi(\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + \lambda^2\int_{\mathcal{S}} g^2(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right] + \mathrm{o}(n^{-4/(d+4)}). \tag{2.17}$$

Theorem 2.5 (Uniform strong consistency)**.** *Assume* (2.1)*. As* $b \to 0$*, we have*

$$\sup_{\boldsymbol{s}\in\mathcal{S}} |f_b(\boldsymbol{s}) - f(\boldsymbol{s})| = \mathcal{O}(b^{1/2}). \tag{2.18}$$

Furthermore, define

$$\mathcal{S}_\delta := \big\{\boldsymbol{s}\in\mathcal{S} : 1 - \|\boldsymbol{s}\|_1 \geq \delta \text{ and } s_i \geq \delta\ \forall i\in[d]\big\}, \quad \delta > 0. \tag{2.19}$$

Then, if $|\log b|^2\, b^{-2d} \leq n$ *as* $n\to\infty$ *and* $b \to 0$*, we have*

$$\sup_{\boldsymbol{s}\in\mathcal{S}_{bd}} |\hat{f}_{n,b}(\boldsymbol{s}) - f(\boldsymbol{s})| = \mathcal{O}\big(|\log b|\, b^{-d}\sqrt{\log n/n}\big) + \mathcal{O}(b^{1/2}), \quad \textit{a.s.} \tag{2.20}$$

In particular, if $|\log b|^2\, b^{-2d} = \mathrm{o}(n/\log n)$*, then* $\sup_{\boldsymbol{s}\in\mathcal{S}_{bd}} |\hat{f}_{n,b}(\boldsymbol{s}) - f(\boldsymbol{s})| \to 0$ *a.s.*

Theorem 2.6 (Asymptotic normality)**.** *Assume* (2.1)*. Let* $\boldsymbol{s}\in\mathrm{Int}(\mathcal{S})$ *be such that* $f(\boldsymbol{s}) > 0$*. If* $n^{1/2}b^{d/4}\to\infty$ *as* $n\to\infty$ *and* $b\to 0$*, then*

$$n^{1/2}b^{d/4}(\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s})) \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(0,\psi(\boldsymbol{s})f(\boldsymbol{s})). \tag{2.21}$$

If we also have $n^{1/2}b^{d/4+1/2}\to 0$ *as* $n\to\infty$ *and* $b\to 0$*, then* (2.18) *of Theorem 2.5 implies*

$$n^{1/2}b^{d/4}(\hat{f}_{n,b}(\boldsymbol{s}) - f(\boldsymbol{s})) \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(0,\psi(\boldsymbol{s})f(\boldsymbol{s})). \tag{2.22}$$

Independently of the above rates for n *and* b*, if we assume* (2.2) *instead and* $n^{2/(d+4)}\, b\to\lambda$ *for some* $\lambda > 0$ *as* $n\to\infty$ *and* $b\to 0$*, then Proposition 2.1 implies*

$$n^{2/(d+4)}(\hat{f}_{n,b}(\boldsymbol{s}) - f(\boldsymbol{s})) \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(\lambda\, g(\boldsymbol{s}), \lambda^{-d/2}\psi(\boldsymbol{s})f(\boldsymbol{s})). \tag{2.23}$$

Remark 2.7. *The rate of convergence for the* d*-dimensional kernel density estimator with i.i.d. data and bandwidth* h *is* $\mathcal{O}(n^{-1/2}h^{-d/2})$ *in Theorem 3.1.15 of Prakasa Rao (1983), whereas our estimator* $\hat{f}_{n,b}$ *converges at a rate of* $\mathcal{O}(n^{-1/2}b^{-d/4})$*. Hence, the relation between the bandwidth of* $\hat{f}_{n,b}$ *and the bandwidth of other multivariate kernel smoothers is* $b\approx h^2$*.*

3. Proof of Proposition 2.1

First, we estimate the expectation and covariances of the random vector

$$\boldsymbol{\xi}_{\boldsymbol{s}} = (\xi_1, \xi_2, \dots, \xi_d) \sim \mathrm{Dirichlet}\big(\tfrac{\boldsymbol{s}}{b} + 1, \tfrac{(1-\|\boldsymbol{s}\|_1)}{b} + 1\big), \quad \boldsymbol{s} \in \mathcal{S}, \ b > 0. \tag{3.1}$$

If $0 < b \le 2^{-1}(d+1)^{-1}$, then, for all $i, j \in [d]$,

$$\begin{aligned} \mathbb{E}[\xi_i] &= \frac{\frac{s_i}{b} + 1}{\frac{1}{b} + d + 1} = \frac{s_i + b}{1 + b(d+1)} \\ &= s_i + b(1 - (d+1)s_i) + \mathcal{O}(b^2), \end{aligned} \tag{3.2}$$

$$\begin{aligned} \mathbb{C}\mathrm{ov}(\xi_i, \xi_j) &= \frac{(\frac{s_i}{b} + 1)((\frac{1}{b} + d + 1)\mathbb{1}_{\{i=j\}} - (\frac{s_j}{b} + 1))}{(\frac{1}{b} + d + 1)^2(\frac{1}{b} + d + 2)} \\ &= \frac{b(s_i + b)(\mathbb{1}_{\{i=j\}} - s_j + b(d+1)\mathbb{1}_{\{i=j\}} - b)}{(1 + b(d+1))^2(1 + b(d+2))} \\ &= b s_i(\mathbb{1}_{\{i=j\}} - s_j) + \mathcal{O}(b^2), \end{aligned} \tag{3.3}$$

$$\begin{aligned} \mathbb{E}[(\xi_i - s_i)(\xi_j - s_j)] &= \mathbb{C}\mathrm{ov}(\xi_i, \xi_j) + (\mathbb{E}[\xi_i] - s_i)(\mathbb{E}[\xi_j] - s_j) \\ &= b s_i(\mathbb{1}_{\{i=j\}} - s_j) + \mathcal{O}(b^2). \end{aligned} \tag{3.4}$$

By a second order mean value theorem, we have

$$\begin{aligned} f(\boldsymbol{\xi}_{\boldsymbol{s}}) - f(\boldsymbol{s}) &= \sum_{i=1}^d \frac{\partial}{\partial x_i} f(\boldsymbol{s})(\xi_i - s_i) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{s})(\xi_i - s_i)(\xi_j - s_j) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \left(\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{\zeta}_{\boldsymbol{s}}) - \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{s})\right)(\xi_i - s_i)(\xi_j - s_j), \end{aligned} \tag{3.5}$$

for some random vector $\boldsymbol{\zeta}_{\boldsymbol{s}} \in \mathcal{S}$ on the line segment joining $\boldsymbol{\xi}_{\boldsymbol{s}}$ and $\boldsymbol{s}$. If we take the expectation in the last equation, and then use (3.2) and (3.4), we get

$$\begin{aligned} &f_b(\boldsymbol{s}) - f(\boldsymbol{s}) - b\left[\sum_{i=1}^d (1 - (d+1)s_i) \frac{\partial}{\partial x_i} f(\boldsymbol{s}) + \frac{1}{2} \sum_{i,j=1}^d s_i(\mathbb{1}_{\{i=j\}} - s_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{s})\right] \\ &\le \frac{1}{2} \sum_{i,j=1}^d \mathbb{E}\left[\left|\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{\zeta}_{\boldsymbol{s}}) - \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{s})\right| |\xi_i - s_i| |\xi_j - s_j| \cdot \mathbb{1}_{\{\|\boldsymbol{\xi}_{\boldsymbol{s}} - \boldsymbol{s}\|_1 \le \delta_{\varepsilon,d}\}}\right] \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \mathbb{E}\left[\left|\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{\zeta}_{\boldsymbol{s}}) - \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{s})\right| |\xi_i - s_i| |\xi_j - s_j| \cdot \mathbb{1}_{\{\|\boldsymbol{\xi}_{\boldsymbol{s}} - \boldsymbol{s}\|_1 > \delta_{\varepsilon,d}\}}\right] \\ &=: \Delta_1 + \Delta_2 \end{aligned} \tag{3.6}$$

where, for any given $\varepsilon > 0$, the real number $\delta_{\varepsilon,d} \in (0,1]$ is such that $\|\boldsymbol{s}' - \boldsymbol{s}\|_1 \le \delta_{d,\varepsilon}$ implies $|\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{s}') - \frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{s})| < \varepsilon$, uniformly for $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$.[2] The uniform continuity of $\big(\frac{\partial^2}{\partial x_i \partial x_j} f\big)_{i,j=1}^d$, the fact that $\|\boldsymbol{\zeta}_{\boldsymbol{s}} - \boldsymbol{s}\|_1 \le \|\boldsymbol{\xi}_{\boldsymbol{s}} - \boldsymbol{s}\|_1$, the Cauchy-Schwarz inequality and (3.4), yield

$$\Delta_1 \le \frac{1}{2} \sum_{i,j=1}^d \varepsilon \cdot \sqrt{\mathbb{E}\big[|\xi_i - s_i|^2\big]} \sqrt{\mathbb{E}\big[|\xi_j - s_j|^2\big]} = \varepsilon \cdot \mathcal{O}(b). \tag{3.7}$$

[2] We know that such a number exists because the second order partial derivatives of f are uniformly continuous on $\mathcal{S}$ by assumption (2.2).

Since $\left(\frac{\partial^2}{\partial x_i \partial x_j} f\right)_{i,j=1}^d$ are uniformly continuous, they are in particular bounded (say by $M_d > 0$). Furthermore, $\{\|\boldsymbol{\xi}_{\boldsymbol{s}} - \boldsymbol{s}\|_1 > \delta_{\varepsilon,d}\}$ implies that at least one component of $(\xi_k - s_k)_{k=1}^d$ is larger than $\delta_{\varepsilon,d}/d$, so a union bound over k followed by d concentration bounds for the beta distribution[3] (see e.g. Lemma A.1 when $d = 1$) yield

$$\begin{aligned} \Delta_2 &\leq \frac{1}{2} \sum_{i,j=1}^d 2M_d \cdot 1^2 \cdot \sum_{k=1}^d \mathbb{P}\big(|\xi_k - s_k| \geq \delta_{\varepsilon,d}/d\big) \\ &\leq d^3\, M_d \cdot 2 \exp\Big(-\frac{(\delta_{\varepsilon,d}/d)^2}{2 \cdot \frac{1}{4}(\frac{1}{b} + d + 2)^{-1}}\Big). \end{aligned} \tag{3.8}$$

If we choose a sequence $\varepsilon = \varepsilon(b) \searrow 0$ that goes to 0 slowly enough that $1 \geq \delta_{\varepsilon,d} > d\sqrt{b \cdot |\log b|}$, then $\Delta_1 + \Delta_2$ in (3.6) is $\mathrm{o}(b)$ by (3.7) and (3.8). This ends the proof of Proposition 2.1.

4. Proof of Theorem 2.2

The expression for the bias is a trivial consequence of Proposition 2.1. In order to compute the asymptotics of the variance, we only assume that f is Lipschitz on $\mathcal{S}$. First, note that

$$\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s}) = \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}), \tag{4.1}$$

where the random variables

$$Y_{i,b}(\boldsymbol{s}) := K_{\frac{\boldsymbol{s}}{b}+1, \frac{1-\|\boldsymbol{s}\|_1}{b}+1}(\boldsymbol{X}_i) - f_b(\boldsymbol{s}), \quad 1 \leq i \leq n, \quad \text{are i.i.d.} \tag{4.2}$$

Hence, if $\boldsymbol{\gamma}_{\boldsymbol{s}} \sim \mathrm{Dirichlet}(2\boldsymbol{s}/b + 1, 2(1 - \|\boldsymbol{s}\|_1)/b + 1)$, then

$$\begin{aligned} \mathbb{V}\mathrm{ar}[\hat{f}_{n,b}(\boldsymbol{s})] &= n^{-1} \mathbb{E}\big[K_{\boldsymbol{s}/b+1,(1-\|\boldsymbol{s}\|_1)/b+1}(\boldsymbol{X})^2\big] - n^{-1}\big(f_b(\boldsymbol{s})\big)^2 \\ &= n^{-1} A_b(\boldsymbol{s})\, \mathbb{E}[f(\boldsymbol{\gamma}_{\boldsymbol{s}})] - \mathcal{O}(n^{-1}) \\ &= n^{-1} A_b(\boldsymbol{s})\,(f(\boldsymbol{s}) + \mathcal{O}(b^{1/2})) - \mathcal{O}(n^{-1}), \end{aligned} \tag{4.3}$$

where

$$A_b(\boldsymbol{s}) := \frac{\Gamma(2(1 - \|\boldsymbol{s}\|_1)/b + 1) \prod_{i=1}^d \Gamma(2s_i/b + 1)}{\Gamma^2((1 - \|\boldsymbol{s}\|_1)/b + 1) \prod_{i=1}^d \Gamma^2(s_i/b + 1)} \cdot \frac{\Gamma^2(1/b + d + 1)}{\Gamma(2/b + d + 1)}, \tag{4.4}$$

and where the last line in (4.3) follows from the Lipschitz continuity of f, the Cauchy-Schwarz inequality and the analogue of (3.4) for $\boldsymbol{\gamma}_{\boldsymbol{s}}$:

$$\begin{aligned} \mathbb{E}[f(\boldsymbol{\gamma}_{\boldsymbol{s}})] - f(\boldsymbol{s}) &= \sum_{i=1}^d \mathcal{O}\Big(\mathbb{E}\big[|\gamma_i - s_i|\big]\Big) \\ &\leq \sum_{i=1}^d \mathcal{O}\Big(\sqrt{\mathbb{E}\big[|\gamma_i - s_i|^2\big]}\Big) = \mathcal{O}(b^{1/2}). \end{aligned} \tag{4.5}$$

The conclusion of Theorem 2.2 follows from (4.3) and Lemma 4.1 below.

[3]The k^{th} component of a Dirichlet$(\boldsymbol{\alpha}, \beta)$ random vector has a Beta$(\alpha_k, \|\boldsymbol{\alpha}\|_1 + \beta - \alpha_k)$ distribution.

Lemma 4.1. *As $b \to 0$, we have, uniformly for $\boldsymbol{s} \in \mathcal{S}$,*

$$0 < A_b(\boldsymbol{s}) \leq \frac{b^{(d+1)/2}\,(1/b+d)^{d+1/2}}{(4\pi)^{d/2}\sqrt{s_1 s_2 \dots s_d(1-\|\boldsymbol{s}\|_1)}}(1+\mathcal{O}(b)). \tag{4.6}$$

Furthermore, for any $\emptyset \neq \mathcal{J} \subseteq [d]$, and any $\boldsymbol{\kappa} \in (0,\infty)^d$,

$$A_b(\boldsymbol{s}) = \begin{cases} b^{-d/2}\,\psi(\boldsymbol{s})(1+\mathcal{O}_{\boldsymbol{s}}(b)), & \text{if } s_i/b \to \kappa_i \ \forall i \in [d] \text{ and} \\ & (1-\|\boldsymbol{s}\|_1)/b \to \infty, \\ b^{-(d+|\mathcal{J}|)/2} \prod_{i\in\mathcal{J}} \frac{\Gamma(2\kappa+1)}{2^{2\kappa_i+1}\Gamma^2(\kappa+1)} \cdot \psi_{\mathcal{J}}(\boldsymbol{s})(1+\mathcal{O}_{\boldsymbol{\kappa},\boldsymbol{s}}(b)), & \text{if } s_i/b \to \kappa_i \ \forall i \in \mathcal{J} \text{ and} \\ & s_i/b \to \infty \ \forall i \in [d]\backslash\mathcal{J} \text{ and} \\ & (1-\|\boldsymbol{s}\|_1)/b \to \infty, \end{cases} \tag{4.7}$$

where ψ and $\psi_{\mathcal{J}}$ are defined as in (2.8).

Proof. If we denote

$$S_b(\boldsymbol{s}) := \frac{R^2((1-\|\boldsymbol{s}\|_1)/b+1)\prod_{i=1}^d R^2(s_i/b+1)}{R(2(1-\|\boldsymbol{s}\|_1)/b+1)\prod_{i=1}^d R(2s_i/b+1)} \cdot \frac{R(2/b+d+1)}{R^2(1/b+d+1)}, \tag{4.8}$$

where

$$R(z) := \frac{\sqrt{2\pi}e^{-z}z^{z+1/2}}{\Gamma(z+1)}, \quad z \geq 0, \tag{4.9}$$

then, for all $\boldsymbol{s} \in \mathrm{Int}(\mathcal{S})$, we have

$$\begin{aligned} A_b(\boldsymbol{s}) &= \frac{2^{2(1-\|\boldsymbol{s}\|_1)/b+1/2}\prod_{i=1}^d 2^{2s_i/b+1/2}}{(2\pi)^{(d+1)/2}\sqrt{(1-\|\boldsymbol{s}\|_1)/b}\,\prod_{i=1}^d\sqrt{s_i/b}} \cdot \frac{\sqrt{2\pi}\,e^{-d}(1/b+d)^{2/b+2d+1}}{(2/b+d)^{2/b+d+1/2}} \cdot S_b(\boldsymbol{s}) \\ &= \frac{b^{(d+1)/2}\,(1/b+d)^{d+1/2}}{(4\pi)^{d/2}\sqrt{s_1 s_2 \dots s_d(1-\|\boldsymbol{s}\|_1)}} \cdot \left(\frac{2/b+2d}{2/b+d}\right)^{2/b+d+1/2} e^{-d} \cdot S_b(\boldsymbol{s}). \end{aligned} \tag{4.10}$$

It was shown in Lemma 3 of Brown & Chen (1999) that $R(z) < 1$ for all $z \geq 0$. Together with the fact that $z \mapsto R(z)$ is increasing on $(1,\infty)$,[4] we see that

$$\max_{\boldsymbol{s}\in\mathcal{S}} S_b(\boldsymbol{s}) \leq \frac{R(2/b+d+1)}{R^2(1/b+d+1)} \to 1, \quad \text{as } b \to 0, \tag{4.11}$$

from which (4.6) follows.

To prove the second claim of the lemma, note that $S_b(\boldsymbol{s}) = 1 + \mathcal{O}_{\boldsymbol{s}}(b)$ by Stirling's formula, so as $s_i/b \to \kappa_i \ \forall i \in [d]$ and $(1-\|\boldsymbol{s}\|_1)/b \to \infty$, we have, from (4.10),

$$\begin{aligned} A_b(\boldsymbol{s}) &= \frac{b^{-d/2}(1+\mathcal{O}(b))}{(4\pi)^{d/2}\sqrt{s_1 s_2 \dots s_d(1-\|\boldsymbol{s}\|_1)}} \cdot \left(1+\frac{d}{2/b+d}\right)^{2/b+d} e^{-d} \cdot (1+\mathcal{O}_{\boldsymbol{s}}(b)) \\ &= \frac{b^{-d/2}(1+\mathcal{O}_{\boldsymbol{s}}(b))}{(4\pi)^{d/2}\sqrt{s_1 s_2 \dots s_d(1-\|\boldsymbol{s}\|_1)}}. \end{aligned} \tag{4.12}$$

[4] By the standard relation $(\Gamma'/\Gamma)(z+1) = 1/z + (\Gamma'/\Gamma)(z)$ and Lemma 2 in Minc & Sathre (1964), we have

$$\frac{\mathrm{d}}{\mathrm{d}z}\log R(z) = \log z + \frac{1}{2z} - \frac{\Gamma'(z+1)}{\Gamma(z+1)} = \log z - \frac{1}{2z} - \frac{\Gamma'(z)}{\Gamma(z)} > 0, \quad \text{for all } z > 1,$$

which means that $z \mapsto R(z)$ is increasing on $(1,\infty)$.

Next, let $\emptyset \neq \mathcal{J} \subseteq [d]$ and $\boldsymbol{\kappa} \in (0,\infty)^d$. If $s_i/b \to \kappa_i$ for all $i \in \mathcal{J}$, $s_i/b \to \infty$ for all $i \in [d]\backslash\mathcal{J}$, and $(1-\|\boldsymbol{s}\|_1)/b \to \infty$, then, from (4.4),

$$\begin{aligned} A_b(\boldsymbol{s}) &= \prod_{i\in\mathcal{J}} \frac{\Gamma(2\kappa+1)}{\Gamma^2(\kappa+1)}(1+\mathcal{O}_{\boldsymbol{\kappa},\boldsymbol{s}}(b)) \cdot \frac{2^{2(1-\|\boldsymbol{s}\|_1)/b+1/2}\prod_{i\in[d]\backslash\mathcal{J}} 2^{2s_i/b+1/2}}{(2\pi)^{(d-|\mathcal{J}|+1)/2}\sqrt{(1-\|\boldsymbol{s}\|_1)/b}\,\prod_{i\in[d]\backslash\mathcal{J}}\sqrt{s_i/b}} \\ &\qquad \cdot \frac{\sqrt{2\pi}\, e^{-d}(1/b+d)^{2/b+2d+1}}{(2/b+d)^{2/b+d+1/2}} \cdot S_b^{\mathcal{J}}(\boldsymbol{s}) \\ &= \prod_{i\in\mathcal{J}} \frac{\Gamma(2\kappa+1)}{\Gamma^2(\kappa+1)}(1+\mathcal{O}_{\boldsymbol{\kappa},\boldsymbol{s}}(b)) \cdot \frac{b^{(d-|\mathcal{J}|+1)/2}\,(1/b+d)^{d+1/2}}{2^{d/2}\prod_{i\in\mathcal{J}} 2^{2\kappa_i+1/2}(2\pi)^{(d-|\mathcal{J}|)/2}\sqrt{(1-\|\boldsymbol{s}\|_1)\prod_{i\in[d]\backslash\mathcal{J}} s_i}} \\ &\qquad \cdot \left(\frac{2/b+2d}{2/b+d}\right)^{2/b+d+1/2} e^{-d}\cdot S_b^{\mathcal{J}}(\boldsymbol{s}), \end{aligned} \tag{4.13}$$

where

$$S_b^{\mathcal{J}}(\boldsymbol{s}) := \frac{R^2((1-\|\boldsymbol{s}\|_1)/b+1)\prod_{i\in[d]\backslash\mathcal{J}} R^2(s_i/b+1)}{R(2(1-\|\boldsymbol{s}\|_1)/b+1)\prod_{i\in[d]\backslash\mathcal{J}} R(2s_i/b+1)} \cdot \frac{R(2/b+d+1)}{R^2(1/b+d+1)}. \tag{4.14}$$

Similarly to (4.12), Stirling's formula implies

$$A_b(\boldsymbol{s}) = \prod_{i\in\mathcal{J}} \frac{\Gamma(2\kappa+1)}{2^{2\kappa_i+1}\Gamma^2(\kappa+1)} \cdot \frac{b^{-(d+|\mathcal{J}|)/2}(1+\mathcal{O}_{\boldsymbol{\kappa},\boldsymbol{s}}(b))}{(4\pi)^{(d-|\mathcal{J}|)/2}\sqrt{(1-\|\boldsymbol{s}\|_1)\prod_{i\in[d]\backslash\mathcal{J}} s_i}}. \tag{4.15}$$

which concludes the proof of Lemma 4.1 and Theorem 2.2. ☐

5. Proof of Theorem 2.4

Proof. By the bound (4.6), the fact that f is uniformly bounded (it is continuous on $\mathcal{S}$), the almost-everywhere convergence in (4.12), and the dominated convergence theorem, we have

$$b^{d/2}\int_{\mathcal{S}} A_b(\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{s} = \int_{\mathcal{S}} \psi(\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + \mathrm{o}(1). \tag{5.1}$$

The expression for the variance in (4.3), and the expression for the bias in (2.7), yield

$$\begin{aligned} \mathrm{MISE}[\hat{f}_{n,b}] &= \int_{\mathcal{S}} \mathbb{V}\mathrm{ar}(\hat{f}_{n,b}(\boldsymbol{s})) + \mathbb{B}\mathrm{ias}[\hat{f}_{n,b}(\boldsymbol{s})]^2\mathrm{d}\boldsymbol{s} \\ &= n^{-1}b^{-d/2}\int_{\mathcal{S}} \psi(\boldsymbol{s})f(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + b^2\int_{\mathcal{S}} g^2(\boldsymbol{s})\mathrm{d}\boldsymbol{s} + \mathrm{o}(n^{-1}b^{-d/2}) + \mathrm{o}(b^2). \end{aligned} \tag{5.2}$$

This ends the proof. ☐

6. Proof of Theorem 2.5

This is the most technical proof, so here is the idea. The first three lemmas below bound, uniformly, the Dirichlet density (Lemma 6.1), the partial derivatives of the Dirichlet density with respect to each of its $(d+1)$ parameters (Lemma 6.2), and then the absolute difference of densities (pointwise and under expectations) that have different parameters (Lemma 6.3). This is then used to show continuity estimates for the random field $\boldsymbol{s} \mapsto Y_{i,b}(\boldsymbol{s})$ from (4.2) (Proposition 6.4), meaning that we get a control on the probability that $Y_{i,b}(\boldsymbol{s})$ and $Y_{i,b}(\boldsymbol{s}')$ are too far apart when $\boldsymbol{s}$ and $\boldsymbol{s}'$ are close. From this, we easily deduce large deviation bounds for the supremum of $Y_{i,b}(\boldsymbol{s})$ over points $\boldsymbol{s}'$ that are inside a small hypercube of width $2b$ centered at $\boldsymbol{s}$ (Corollary 6.5). Since $\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s}) = \frac{1}{n}\sum_{i=1}^n Y_{i,b}(\boldsymbol{s})$, we can estimate tail probabilities for the global supremum of $|\hat{f}_{n,b} - f_b|$ by a union bound over suprema on the small hypercubes and apply a large deviation bound from Corollary 6.5 for each one of them.

The first lemma bounds the density of the Dirichlet$(\boldsymbol{\alpha}, \beta)$ distribution from (1.2).

Lemma 6.1. *If* $\alpha_1, \dots, \alpha_d, \beta \geq 2$, *then*

$$\sup_{\boldsymbol{s}\in\mathcal{S}} K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s}) \leq \sqrt{\frac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}}\,(\|\boldsymbol{\alpha}\|_1 + \beta - d - 1)^d. \tag{6.1}$$

Proof. The Dirichlet density is well-known to be maximized at $\boldsymbol{s}^\star = \frac{\boldsymbol{\alpha}-1}{\|\boldsymbol{\alpha}\|_1+\beta-d-1}$. At this point, we have

$$K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s}^\star) = \frac{\Gamma(\|\boldsymbol{\alpha}\|_1 + \beta)}{\Gamma(\beta)\prod_{i=1}^d \Gamma(\alpha_i)} \cdot \frac{(\beta-1)^{\beta-1}\prod_{i=1}^d(\alpha_i-1)^{\alpha_i-1}}{(\|\boldsymbol{\alpha}\|_1 + \beta - d - 1)^{\|\boldsymbol{\alpha}\|_1+\beta-d-1}}. \tag{6.2}$$

From Theorem 2.2 in Batir (2017), we know that

$$\sqrt{2\pi}e^{-y+1}(y-1)^{y-1+\frac{1}{2}} \leq \Gamma(y) \leq \tfrac{7}{5}\cdot\sqrt{2\pi}e^{-y+1}(y-1)^{y-1+\frac{1}{2}}, \quad \text{for all } y \geq 2, \tag{6.3}$$

so (6.2) is

$$\begin{aligned}
&\leq \frac{\frac{7}{5}\sqrt{2\pi}e^{-\|\boldsymbol{\alpha}\|_1-\beta+1}(\|\boldsymbol{\alpha}\|_1+\beta-1)^{\|\boldsymbol{\alpha}\|_1+\beta-1+\frac{1}{2}}}{\sqrt{2\pi}e^{-\beta+1}(\beta-1)^{\beta-1+\frac{1}{2}}\prod_{i=1}^d\sqrt{2\pi}e^{-\alpha_i+1}(\alpha_i-1)^{\alpha_i-1+\frac{1}{2}}} \cdot \frac{(\beta-1)^{\beta-1}\prod_{i=1}^d(\alpha_i-1)^{\alpha_i-1}}{(\|\boldsymbol{\alpha}\|_1+\beta-d-1)^{\|\boldsymbol{\alpha}\|_1+\beta-d-1}} \\
&= \tfrac{7}{5}\,(2\pi)^{-d/2}\cdot e^{-d}\left(1 - \frac{d}{\|\boldsymbol{\alpha}\|_1+\beta-1}\right)^{-(\|\boldsymbol{\alpha}\|_1+\beta-1)} \cdot \sqrt{\frac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}} \cdot (\|\boldsymbol{\alpha}\|_1 + \beta - d - 1)^d \\
&\leq \tfrac{7}{5}\,(2\pi)^{-d/2}\cdot e^{\frac{2}{5}d}\cdot \sqrt{\frac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}} \cdot (\|\boldsymbol{\alpha}\|_1 + \beta - d - 1)^d \\
&\leq \sqrt{\frac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}} \cdot (\|\boldsymbol{\alpha}\|_1 + \beta - d - 1)^d,
\end{aligned} \tag{6.4}$$

where we used our assumption that $\alpha_1, \dots, \alpha_d, \beta \geq 2$ and the fact that $(1-y)^{-1} \leq e^{\frac{7}{5}y}$ for $y \in [0, 1/2]$ to obtain the second inequality. The conclusion follows. ☐

In the next lemma, we bound the partial derivatives of the Dirichlet$(\boldsymbol{\alpha}, \beta)$ density function with respect to its parameters.

Lemma 6.2. *If* $\alpha_1, \dots, \alpha_d, \beta \geq 2$, *then*

$$\begin{aligned}
\sup_{\boldsymbol{s}\in\mathcal{S}} \left|\frac{\mathrm{d}}{\mathrm{d}\alpha_j} K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s})\right| &\leq \Big\{2\,|\log(\|\boldsymbol{\alpha}\|_1+\beta)| + |\log s_j|\Big\} \\
&\qquad \cdot \sqrt{\frac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}}\,(\|\boldsymbol{\alpha}\|_1 + \beta - d - 1)^d,
\end{aligned} \tag{6.5}$$

$$\begin{aligned}
\sup_{\boldsymbol{s}\in\mathcal{S}} \left|\frac{\mathrm{d}}{\mathrm{d}\beta} K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s})\right| &\leq \Big\{2\,|\log(\|\boldsymbol{\alpha}\|_1+\beta)| + |\log(1-\|\boldsymbol{s}\|_1)|\Big\} \\
&\qquad \cdot \sqrt{\frac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}}\,(\|\boldsymbol{\alpha}\|_1 + \beta - d - 1)^d.
\end{aligned} \tag{6.6}$$

Proof. The digamma function $\psi(z) := \Gamma'(z)/\Gamma(z)$ satisfies $\psi(z) < \log(z)$ for all $z \geq 1$ (see e.g. Lemma 2 in Minc & Sathre (1964)), so we have

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}\alpha_j} K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s})\right| &= \left|\big(\psi(\|\boldsymbol{\alpha}\|_1 + \beta) - \psi(\alpha_j) + \log s_j\big) \, K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s})\right| \\ &\leq \Big\{2\,|\log(\|\boldsymbol{\alpha}\|_1 + \beta)| + |\log s_j|\Big\} \, K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s}). \qquad (6.7)\end{aligned}$$

The conclusion (6.5) follows from Lemma 6.1. The proof of (6.6) is virtually the same. $\square$

As a consequence of Lemma 6.2 and the multivariate mean value theorem, we can control the difference of two Dirichlet densities with different parameters, pointwise and under expectations.

Lemma 6.3. *If $\alpha_1, \dots, \alpha_d, \beta, \alpha_1', \dots, \alpha_d', \beta' \geq 2$, and $\boldsymbol{X}$ is F distributed with a bounded density f supported on $\mathcal{S}$, then*

$$\begin{aligned}&\mathbb{E}\big[|K_{\boldsymbol{\alpha}',\beta'}(\boldsymbol{X}) - K_{\boldsymbol{\alpha},\beta}(\boldsymbol{X})|\big] \\ &\quad\leq C\, d\, \|f\|_\infty \sqrt{\tfrac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}} \, \big(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta') - d - 1\big)^d \qquad (6.8)\\ &\qquad \cdot \log\big(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta')\big) \cdot \|(\boldsymbol{\alpha}', \beta') - (\boldsymbol{\alpha}, \beta)\|_\infty,\end{aligned}$$

where $C > 0$ is a universal constant. Furthermore, if

$$\mathcal{S}_\delta := \big\{\boldsymbol{s} \in \mathcal{S} : 1 - \|\boldsymbol{s}\|_1 \geq \delta \text{ and } s_i \geq \delta \ \forall i \in [d]\big\}, \quad \delta > 0, \qquad (6.9)$$

then we have

$$\begin{aligned}&\max_{\boldsymbol{s} \in \mathcal{S}_\delta} |K_{\boldsymbol{\alpha}',\beta'}(\boldsymbol{s}) - K_{\boldsymbol{\alpha},\beta}(\boldsymbol{s})| \\ &\quad\leq \widetilde{C}\, d\, \|f\|_\infty \log \delta \cdot \sqrt{\tfrac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}} \, \big(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta') - d - 1\big)^d \qquad (6.10)\\ &\qquad \cdot \log\big(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta')\big) \cdot \|(\boldsymbol{\alpha}', \beta') - (\boldsymbol{\alpha}, \beta)\|_\infty,\end{aligned}$$

where $\widetilde{C} > 0$ is another universal constant.

Proof. By the multivariate mean value theorem and Lemma 6.2,

$$\begin{aligned}&\mathbb{E}\big[|K_{\boldsymbol{\alpha}',\beta'}(\boldsymbol{X}) - K_{\boldsymbol{\alpha},\beta}(\boldsymbol{X})|\big] \\ &\quad\leq \|f\|_\infty \sqrt{\tfrac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta-1)\prod_{i=1}^d(\alpha_i - 1)}} \, \big(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta') - d - 1\big)^d \log\big(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta')\big) \\ &\qquad \cdot \|(\boldsymbol{\alpha}', \beta') - (\boldsymbol{\alpha}, \beta)\|_\infty \bigg\{\int_{\mathcal{S}} |\log(1 - \|\boldsymbol{s}\|_1)| \mathrm{d}\boldsymbol{s} + \sum_{j=1}^d \int_{\mathcal{S}} |\log s_j| \mathrm{d}\boldsymbol{s}\bigg\}.\end{aligned} \qquad (6.11)$$

The integrals are bounded by 1. Indeed, if $\widetilde{\mathcal{S}}$ denotes a $(d-1)$-dimensional simplex, then

$$\int_{\mathcal{S}} |\log s_j| \mathrm{d}\boldsymbol{s} = \int_0^1 |\log s_j| \cdot \underbrace{\mathrm{Vol}_{d-1}((1 - s_j)\widetilde{\mathcal{S}})}_{\leq 1} \mathrm{d}s_j \leq \int_0^1 |\log s_j| \mathrm{d}s_j = 1. \qquad (6.12)$$

The second claim is obvious, again by the multivariate mean value theorem and Lemma 6.2. $\square$

Proposition 6.4 (Continuity estimates)**.** *Let $b > 0$ and recall from (4.2) :*

$$Y_{i,b}(\boldsymbol{s}) := K_{\frac{\boldsymbol{s}}{b}+1, \frac{1-\|\boldsymbol{s}\|_1}{b}+1}(\boldsymbol{X}_i) - \mathbb{E}\big[K_{\frac{\boldsymbol{s}}{b}+1, \frac{1-\|\boldsymbol{s}\|_1}{b}+1}(\boldsymbol{X}_i)\big], \quad 1 \leq i \leq n. \qquad (6.13)$$

If $\boldsymbol{s} \in \mathcal{S}_{b(d+1)}$ and $|\log b|^2\, b^{-2d} \leq n$, then we have, for any $h \in \mathbb{R}$ and $a \geq 1$,

$$\mathbb{P}\Big(\sup_{\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d} \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}') \geq h + a, \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}) \leq h \Big) \leq \exp\Big(- c_{f,d} \cdot \frac{n\, b^{2d} a^2}{|\log b|^2} \Big). \tag{6.14}$$

where $c_{f,d} > 0$ depends only on the density f and the dimension d.

Proof. Clearly, the probability in (6.14) is

$$\leq \mathbb{P}\Big(\sup_{\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d} \frac{1}{n} \sum_{i=1}^n \big(Y_{i,b}(\boldsymbol{s}') - Y_{i,b}(\boldsymbol{s}) \big) \geq a \Big). \tag{6.15}$$

The main idea of the proof now is to decompose the supremum with a chaining argument and apply concentration bounds on the increments at each level of the d-dimensional tree. With the notation $\mathcal{H}_k := 2^{-k} \cdot b\, \mathbb{Z}^d$, we have the embedded sequence of lattice points

$$\mathcal{H}_0 \subseteq \mathcal{H}_1 \subseteq \cdots \subseteq \mathcal{H}_k \subseteq \cdots \subseteq \mathbb{R}^d. \tag{6.16}$$

Hence, for $\boldsymbol{s} \in \mathcal{S}_{b(d+1)}$ fixed, and for any $\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d$, let $(\boldsymbol{s}_k)_{k \in \mathbb{N}_0}$ be a sequence that satisfies

$$\boldsymbol{s}_0 = \boldsymbol{s}, \qquad \boldsymbol{s}_k - \boldsymbol{s} \in \mathcal{H}_k \cap [-b,b]^d, \qquad \lim_{k \to \infty} \|\boldsymbol{s}_k - \boldsymbol{s}'\|_\infty = 0, \tag{6.17}$$

and

$$(\boldsymbol{s}_{k+1})_i = (\boldsymbol{s}_k)_i \pm 2^{-k-1} b, \quad \text{for all } i \in [d]. \tag{6.18}$$

Since the map $\boldsymbol{s} \mapsto \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s})$ is almost-surely continuous,

$$\frac{1}{n} \sum_{i=1}^n \big(Y_{i,b}(\boldsymbol{s}') - Y_{i,b}(\boldsymbol{s}) \big) = \sum_{k=0}^\infty \frac{1}{n} \sum_{i=1}^n \big(Y_{i,b}(\boldsymbol{s}_{k+1}) - Y_{i,b}(\boldsymbol{s}_k) \big), \tag{6.19}$$

and since, $\sum_{k=0}^\infty \frac{1}{2(k+1)^2} \leq 1$, we have the inclusion of events,

$$\begin{aligned} &\Big\{ \sup_{\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d} \frac{1}{n} \sum_{i=1}^n \big(Y_{i,b}(\boldsymbol{s}') - Y_{i,b}(\boldsymbol{s}) \big) \geq a \Big\} \\ &\subseteq \bigcup_{k=0}^\infty \bigcup_{\substack{\boldsymbol{s}_1 \in \boldsymbol{s} + \mathcal{H}_k \cap (-b,b)^d \\ (\boldsymbol{s}_2)_i = (\boldsymbol{s}_1)_i \pm 2^{-k-1} b \ \forall i \in [d]}} \Big\{ \frac{1}{n} \sum_{i=1}^n \big(Y_{i,b}(\boldsymbol{s}_2) - Y_{i,b}(\boldsymbol{s}_1) \big) \geq \frac{a}{2(k+1)^2} \Big\}. \end{aligned} \tag{6.20}$$

By a union bound and the fact that $|\mathcal{H}_k \cap (-b,b)^d| \leq 2^{(k+1)d}$, the probability in (6.15) is

$$\leq \sum_{k=0}^\infty 2^{(k+1)d} \cdot 2^d \sup_{\substack{\boldsymbol{s}_1 \in \boldsymbol{s} + \mathcal{H}_k \cap (-b,b)^d \\ (\boldsymbol{s}_2)_i = (\boldsymbol{s}_1)_i \pm 2^{-k-1} b \ \forall i \in [d]}} \mathbb{P}\Big(\frac{1}{n} \sum_{i=1}^n \big(Y_{i,b}(\boldsymbol{s}_2) - Y_{i,b}(\boldsymbol{s}_1) \big) \geq \frac{a}{2(k+1)^2} \Big). \tag{6.21}$$

By Azuma's inequality (see e.g. Theorem 1.3.1 in Steele (1997)), Lemma 6.3 (note that $\boldsymbol{s} \in \mathcal{S}_{b(d+1)}$ and $\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d$ imply $\boldsymbol{s}' \in \mathcal{S}_b$) and (6.13), the above is

$$\begin{aligned} &\leq \sum_{k=0}^\infty 2^{(k+2)d} \cdot 2 \exp\Big(- \frac{na^2}{8(k+1)^4} \cdot \big(C_{f,d}\, b^{-d} \log(b^{-1} + d + 1)\, 2^{-k-1} \big)^{-2} \Big) \\ &\leq \sum_{k=0}^\infty 2^{(k+2)d} \cdot 2 \exp\Big(- \frac{2^{2k-1}}{C_{f,d}^2 \,(k+1)^4} \cdot \frac{n\, b^{2d} a^2}{|\log b|^2} \Big), \end{aligned} \tag{6.22}$$

for some constant $C_{f,d} > 0$ that only depend on f and d. Assuming $|\log b|^2\, b^{-2d} \leq n$, this is

$$\leq \exp\Big(- c_{f,d} \cdot \frac{n\, b^{2d} a^2}{|\log b|^2} \Big), \tag{6.23}$$

for some other constant $c_{f,d} > 0$. This ends the proof of Proposition 6.4. $\square$

Corollary 6.5 (Large deviation estimates)**.** *If $\boldsymbol{s} \in \mathcal{S}_{b(d+1)}$ and $|\log b|^2\, b^{-2d} \leq n$, then we have, for any $a \geq 1$,*

$$\mathbb{P}\bigg(\sup_{\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d} \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}') \geq 2a \bigg) \leq \exp\Big(- c_{f,d} \cdot \frac{n\, b^{2d} a^2}{|\log b|^2} \Big), \tag{6.24}$$

where $c_{f,d} > 0$ depends only on the density f and the dimension d.

Proof. By a union bound, the probability in (6.24) is

$$\begin{aligned} &\leq \mathbb{P}\bigg(\sup_{\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d} \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}') \geq 2a, \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}) \leq a \bigg) \\ &\quad + \mathbb{P}\bigg(\frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}) \geq a \bigg). \end{aligned} \tag{6.25}$$

The first probability is bounded with Proposition 6.4 with $h = a$. We get the same bound on the second probability by applying Azuma's inequality and Lemma 6.3, as we did in (6.22). ☐

We are now ready to prove Theorem 2.5. On the one hand, the Lipschitz continuity of f, Jensen's inequality and (3.4), imply that, uniformly for $\boldsymbol{s} \in \mathcal{S}$,

$$\begin{aligned} f_b(\boldsymbol{s}) - f(\boldsymbol{s}) = \mathbb{E}[f(\boldsymbol{\xi}_{\boldsymbol{s}})] - f(\boldsymbol{s}) &= \sum_{i=1}^d \mathcal{O}\Big(\mathbb{E}\big[|\xi_i - s_i|\big] \Big) \\ &\leq \sum_{i=1}^d \mathcal{O}\Big(\sqrt{\mathbb{E}\big[|\xi_i - s_i|^2\big]} \Big) = \mathcal{O}(b^{1/2}). \end{aligned} \tag{6.26}$$

On the other hand, recall from (4.1) that

$$\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s}) = \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}), \quad \text{where } Y_{i,b}(\boldsymbol{s}) := K_{\frac{\boldsymbol{s}}{b}+1, \frac{1-\|\boldsymbol{s}\|_1}{b}+1}(\boldsymbol{X}_i) - f_b(\boldsymbol{s}). \tag{6.27}$$

By a union bound over the suprema on hypercubes of width $2b$ centered at each $\boldsymbol{s} \in 2b\,\mathbb{Z}^d \cap \mathcal{S}_{b(d+1)}$, and the large deviation estimates in Corollary 6.5 (assuming $|\log b|^2\, b^{-2d} \leq n$), we have

$$\begin{aligned} \mathbb{P}\Big(\sup_{\boldsymbol{s} \in \mathcal{S}_{bd}} |\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s})| > 2a \Big) &\leq \sum_{\boldsymbol{s} \in 2b\,\mathbb{Z}^d \cap \mathcal{S}_{b(d+1)}} \mathbb{P}\Big(\sup_{\boldsymbol{s}' \in \boldsymbol{s} + [-b,b]^d} \frac{1}{n} \sum_{i=1}^n Y_{i,b}(\boldsymbol{s}) > 2a \Big) \\ &\leq b^{-d} \cdot \exp\Big(- c_{f,d} \cdot \frac{n\, b^{2d} (2a)^2}{|\log b|^2} \Big). \end{aligned} \tag{6.28}$$

With the choice $a = (c_{f,d})^{-1/2}\, |\log b|\, b^{-d} \sqrt{\log n / n}$, the right-hand side of (6.28) is equal to $\exp(-4 \log n + d |\log b|)$, which is summable in n,[5] so the Borel-Cantelli lemma yields

$$\sup_{\boldsymbol{s} \in \mathcal{S}_{bd}} |\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s})| = \mathcal{O}\big(|\log b|\, b^{-d} \sqrt{\log n / n} \big), \quad \text{a.s.} \tag{6.29}$$

Together with (6.26), the conclusion follows.

[5]Indeed, our assumption $|\log b|^2\, b^{-2d} \leq n$ implies $d|\log b| \leq \log n$ for example, and n^{-3} is summable.

7. Proof of Theorem 2.6

By (6.27), the asymptotic normality of $n^{1/2}b^{d/4}(\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s}))$ will follow if we verify the Lindeberg condition for double arrays :[6] For every $\varepsilon > 0$,

$$s_b^{-2}\,\mathbb{E}\big[|Y_{1,b}(\boldsymbol{s})|^2\,\mathbb{1}_{\{|Y_{1,b}(\boldsymbol{s})|>\varepsilon n^{1/2}s_b\}}\big] \longrightarrow 0, \quad \text{as } n \to \infty, \tag{7.1}$$

where $s_b^2 := \mathbb{E}\big[|Y_{1,b}(\boldsymbol{s})|^2\big]$ and $b = b(n) \to 0$. From Lemma 6.1, we know that

$$|Y_{1,b}(\boldsymbol{s})| = \mathcal{O}\big(\psi(\boldsymbol{s})\,b^{d/2}\cdot b^{-d}\big) = \mathcal{O}_{\boldsymbol{s}}(b^{-d/2}), \tag{7.2}$$

and we also know that $s_b = b^{-d/4}\sqrt{\psi(\boldsymbol{s})f(\boldsymbol{s})}(1 + \mathrm{o}_{\boldsymbol{s}}(1))$ when f is Lipschitz continuous, by the proof of Theorem 2.2, so

$$\frac{|Y_{1,b}(\boldsymbol{s})|}{n^{1/2}s_b} = \mathcal{O}_{\boldsymbol{s}}(n^{-1/2}\,b^{d/4}\,b^{-d/2}) = \mathcal{O}_{\boldsymbol{s}}(n^{-1/2}b^{-d/4}) \longrightarrow 0, \tag{7.3}$$

whenever $n^{1/2}b^{d/4} \to \infty$ as $n \to \infty$ and $b \to 0$. Under this condition, (7.1) holds and thus

$$n^{1/2}b^{d/4}(\hat{f}_{n,b}(\boldsymbol{s}) - f_b(\boldsymbol{s})) = n^{1/2}b^{d/4}\cdot\frac{1}{n}\sum_{i=1}^{n}Y_{i,m} \stackrel{\mathscr{D}}{\longrightarrow} \mathcal{N}(0,\psi(\boldsymbol{s})f(\boldsymbol{s})). \tag{7.4}$$

This completes the proof of Theorem 2.6.

A. Supplemental material

The sub-Gaussianity property of the Dirichlet distribution allows us to get very useful concentration bounds. The following lemma is used in the proof of Proposition 2.1.

Lemma A.1 (Concentration bounds for the Dirichlet distribution)**.** *Let* $\boldsymbol{D} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha},\beta)$ *for some* $\alpha_1,\dots,\alpha_d,\beta > 0$*. There is a variance parameter* $0 < \sigma_{\mathrm{opt}}^2(\boldsymbol{\alpha},\beta) \leq (4(\|\boldsymbol{\alpha}\|_1 + \beta + 1))^{-1}$ *such that*

$$\mathbb{P}\big(|(\boldsymbol{D} - \mathbb{E}[\boldsymbol{D}])_i| \geq t_i \ \forall i \in [d]\big) \leq 2^d \exp\Big(-\frac{\|\boldsymbol{t}\|_2^2}{2\sigma_{\mathrm{opt}}^2(\boldsymbol{\alpha},\beta)}\Big), \tag{A.1}$$

for all $\boldsymbol{t} \in (0,\infty)^d$*.*

Proof. By Chernoff's inequality and the sub-Gaussianity of the Dirichlet distribution, shown in Theorem 3.3 of Marchal & Arbel (2017), we have, for all $\boldsymbol{\lambda} \in (0,\infty)^d$,

$$\mathbb{P}(\boldsymbol{D} - \mathbb{E}[\boldsymbol{D}] \geq \boldsymbol{t}) \leq \mathbb{E}\big[e^{\boldsymbol{\lambda}^\top(\boldsymbol{D} - \mathbb{E}[\boldsymbol{D}])}\big]e^{-\boldsymbol{\lambda}^\top \boldsymbol{t}} \leq e^{-\frac{\|\boldsymbol{\lambda}\|_2^2\sigma_{\mathrm{opt}}^2(\boldsymbol{\alpha},\beta)}{2} - \boldsymbol{\lambda}^\top \boldsymbol{t}}, \tag{A.2}$$

for some $0 < \sigma_{\mathrm{opt}}^2(\boldsymbol{\alpha},\beta) \leq (4(\|\boldsymbol{\alpha}\|_1 + \beta + 1))^{-1}$. (The upper bound on $\sigma_{\mathrm{opt}}^2(\boldsymbol{\alpha},\beta)$ is stated in Theorem 2.1 of Marchal & Arbel (2017).) If we take the optimal vector $\boldsymbol{\lambda}^\star = \boldsymbol{t}/\sigma_{\mathrm{opt}}^2(\boldsymbol{\alpha},\beta)$, the right-hand side of (A.2) is $\leq \exp(-\|\boldsymbol{t}\|_2^2/(2\sigma_{\mathrm{opt}}^2(\boldsymbol{\alpha},\beta)))$. We get the same bound on any probability of the form

$$\mathbb{P}\big((\boldsymbol{D} - \mathbb{E}[\boldsymbol{D}])_i \leq -t_i \ \forall i \in \mathcal{J}, (\boldsymbol{D} - \mathbb{E}[\boldsymbol{D}])_i \geq t_i \ \forall [d]\backslash\mathcal{J}\big), \quad \text{for } \mathcal{J} \subseteq [d], \tag{A.3}$$

simply by rerunning the above argument and multiplying the components $i \in \mathcal{J}$ of $\boldsymbol{\lambda}^\star$ by -1. Since there are 2^d possible subsets of $[d]$, the conclusion (A.1) follows from a union bound. $\square$

[6] See e.g. Section 1.9.3. in Serfling (1980).

References

Abramowitz, M., & Stegun, I. A. 1964. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. National Bureau of Standards Applied Mathematics Series, vol. 55. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. MR0167642.

Babu, G. J., & Chaubey, Y. P. 2006. Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors. *Statist. Probab. Lett.*, **76**(9), 959–969. MR2270097.

Babu, G. J., Canty, A. J., & Chaubey, Y. P. 2002. Application of Bernstein polynomials for smooth estimation of a distribution and density function. *J. Statist. Plann. Inference*, **105**(2), 377–392. MR1910059.

Barrientos, A. F., Jara, A., & Quintana, F. A. 2015. Bayesian density estimation for compositional data using random Bernstein polynomials. *J. Statist. Plann. Inference*, **166**, 116–125. MR3390138.

Barrientos, A. F., Jara, A., & Quintana, F. A. 2017. Fully nonparametric regression for bounded data using dependent Bernstein polynomials. *J. Amer. Statist. Assoc.*, **112**(518).

Batir, N. 2017. Bounds for the gamma function. *Results Math.*, **72**(no. 1-2), 865–874. MR3684463.

Belalia, M. 2016. On the asymptotic properties of the Bernstein estimator of the multivariate distribution function. *Statist. Probab. Lett.*, **110**, 249–256. MR3474765.

Belalia, M., Bouezmarni, T., & Leblanc, A. 2017a. Smooth conditional distribution estimators using Bernstein polynomials. *Comput. Statist. Data Anal.*, **111**, 166–182. MR3630225.

Belalia, M., Bouezmarni, T., Lemyre, F. C., & Taamouti, A. 2017b. Testing independence based on Bernstein empirical copula and copula density. *J. Nonparametr. Stat.*, **29**(2), 346–380. MR3635017.

Belalia, M., Bouezmarni, T., & Leblanc, A. 2019. Bernstein conditional density estimation with application to conditional distribution and regression functions. *J. Korean Statist. Soc.*, **48**(3), 356–383. MR3983257.

Bouezmarni, T., & Rolin, J.-M. 2003. Consistency of the beta kernel density function estimator. *Canad. J. Statist.*, **31**(1), 89–98. MR1985506.

Bouezmarni, T., & Rolin, J.-M. 2007. Bernstein estimator for unbounded density function. *J. Nonparametr. Stat.*, **19**(3), 145–161. MR2351744.

Bouezmarni, T., & Scaillet, O. 2005. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*, **21**(2), 390–412. MR2179543.

Bouezmarni, T., M., Mesfioui, & Rolin, J. M. 2007. L_1-rate of convergence of smoothed histogram. *Statist. Probab. Lett.*, **77**(14), 1497–1504. MR2395599.

Bouezmarni, T., Rombouts, J. V. K., & Taamouti, A. 2010. Asymptotic properties of the Bernstein density copula estimator for α-mixing data. *J. Multivariate Anal.*, **101**(1), 1–10. MR2557614.

Bouezmarni, T., El Ghouch, A., & Taamouti, A. 2013. Bernstein estimator for unbounded copula densities. *Stat. Risk Model.*, **30**(4), 343–360. MR3143795.

Brown, B. M., & Chen, S. X. 1999. Beta-Bernstein smoothing for regression curves with compact support. *Scand. J. Statist.*, **26**(1), 47–59. MR1685301.

Chen, S. X. 1999. Beta kernel estimators for density functions. *Comput. Statist. Data Anal.*, **31**(2), 131–145. MR1718494.

Chen, S. X. 2000a. Beta kernel smoothers for regression curves. *Statist. Sinica*, **10**(1), 73–91. MR1742101.

Chen, S. X. 2000b. Probability density function estimation using gamma kernels. *Ann. Inst. Statist. Math*, **52**(3), 471–480. MR1794247.

Curtis, S. M., & Ghosh, S. K. 2011. A variable selection approach to monotonic regression with Bernstein polynomials. *J. Appl. Stat.*, **38**(5), 961–976. MR2782409.

Gawronski, W. 1985. Strong laws for density estimators of Bernstein type. *Period. Math. Hungar*, **16**(1), 23–43. MR0791719.

Gawronski, W., & Stadtmüller, U. 1980. On density estimation by means of Poisson's distribution. *Scand. J. Statist.*, **7**(2), 90–94. MR0574548.

Gawronski, W., & Stadtmüller, U. 1981. Smoothing histograms by means of lattice and continuous distributions. *Metrika*, **28**(3), 155–164. MR0638651.

Gawronski, W., & Stadtmüller, U. 1984. Linear combinations of iterated generalized Bernstein functions with an application to density estimation. *Acta Sci. Math.*, **47**(1-2), 205–221. MR0755576.

Ghosal, S. 2001. Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.*, **29**(5), 1264–1280. MR1873330.

Guan, Z. 2016. Efficient and robust density estimation using Bernstein type polynomials. *J. Nonparametr. Stat.*, **28**(2), 250–271. MR3488598.

Igarashi, G., & Kakizawa, Y. 2014. On improving convergence rate of Bernstein polynomial density estimator. *J. Nonparametr. Stat.*, **26**(1), 61–84. MR3174309.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2012. Large sample behavior of the Bernstein copula estimator. *J. Statist. Plann. Inference*, **142**(5), 1189–1197. MR2879763.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2014. A note on the asymptotic behavior of the Bernstein estimator of the copula density. *J. Multivariate Anal.*, **124**, 480–487. MR3147339.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2016. Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals. *TEST*, **25**(2), 351–374. MR3493523.

Janssen, P., Swanepoel, J., & Veraverbeke, N. 2017. Smooth copula-based estimation of the conditional density function with a single covariate. *J. Multivariate Anal.*, **159**, 39–48. MR3668546.

Kakizawa, Y. 2004. Bernstein polynomial probability density estimation. *J. Nonparametr. Stat.*, **16**(5), 709–729. MR2068610.

Kakizawa, Y. 2006. Bernstein polynomial estimation of a spectral density. *J. Time Ser. Anal.*, **27**(2), 253–287. MR2235846.

Kakizawa, Y. 2011. A note on generalized Bernstein polynomial density estimators. *Stat. Methodol.*, **8**(2), 136–153. MR2769276.

Lange, K. 1995. Applications of the Dirichlet distribution to forensic match probabilities. *Genetica*, **96**(1-2), 107–117. doi:10.1007/BF01441156.

Leblanc, A. 2009. Chung-Smirnov property for Bernstein estimators of distribution functions. *J. Nonparametr. Stat.*, **21**(2), 133–142. MR2488150.

Leblanc, A. 2010. A bias-reduced approach to density estimation using Bernstein polynomials. *J. Nonparametr. Stat.*, **22**(3-4), 459–475. MR2662607.

Leblanc, A. 2012a. On estimating distribution functions using Bernstein polynomials. *Ann. Inst. Statist. Math.*, **64**(5), 919–943. MR2960952.

Leblanc, A. 2012b. On the boundary properties of Bernstein polynomial estimators of density and distribution functions. *J. Statist. Plann. Inference*, **142**(10), 2762–2778. MR2925964.

Lorentz, G. G. 1986. *Bernstein polynomials*. Second edition. Chelsea Publishing Co., New York. MR0864976.

Lu, L. 2015. On the uniform consistency of the Bernstein density estimator. *Statist. Probab. Lett.*, **107**, 52–61. MR3412755.

Marchal, O., & Arbel, J. 2017. On the sub-Gaussianity of the beta and Dirichlet distributions. *Electron. Commun. Probab.*, **22**(Paper No. 54), 14 pp. MR3718704.

Minc, H., & Sathre, L. 1964. Some inequalities involving $(r!)^{1/r}$. *Proc. Edinburgh Math. Soc.*, **14**(1), 41–46. MR0162751.

Ouimet, F. 2018. Complete monotonicity of multinomial probabilities and its application to Bernstein estimators on the simplex. *J. Math. Anal. Appl.*, **466**(2), 1609–1617. MR3825458.

Ouimet, F. 2019. *Extremes of log-correlated random fields and the Riemann-zeta function, and some asymptotic results for various estimators in statistics*. PhD thesis, Université de Montréal. `http://hdl.handle.net/1866/22667`.

Ouimet, F. 2020. A precise local limit theorem for the multinomial distribution. *Preprint*, 1–7. arXiv:2001.08512.

Petrone, S. 1999a. Bayesian density estimation using Bernstein polynomials. *Canad. J. Statist.*, **27**(1), 105–126. MR1703623.

Petrone, S. 1999b. Random Bernstein polynomials. *Scand. J. Statist.*, **26**(3), 373–393. MR1712051.

Petrone, S., & Wasserman, L. 2002. Consistency of Bernstein polynomial posteriors. *J. Roy. Statist. Soc. Ser. B*, **64**(1), 79–100. MR1881846.

Prakasa Rao, B. L. S. 1983. *Nonparametric functional estimation*. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York. MR0740865.

Prakasa Rao, B. L. S. 2005. Estimation of distribution and density functions by generalized Bernstein polynomials. *Indian J. Pure Appl. Math.*, **36**(2), 63–88. MR2153833.

Sancetta, A. 2007. Nonparametric estimation of distributions with given marginals via Bernstein-Kantorovich polynomials: L_1 and pointwise convergence theory. *J. Multivariate Anal.*, **98**(7), 1376–1390. MR2364124.

Sancetta, A., & Satchell, S. 2004. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, **20**(3), 535–562. MR2061727.

Serfling, R. J. 1980. *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. MR0595165.

Stadtmüller, U. 1983. Asymptotic distributions of smoothed histograms. *Metrika*, **30**(3), 145–158. MR0726014.

Stadtmüller, U. 1986. Asymptotic properties of nonparametric curve estimates. *Period. Math. Hungar.*, **17**(2), 83–108. MR0858109.

Steele, M. 1997. *Probability theory and combinatorial optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 69. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. MR1422018.

Tenbusch, A. 1994. Two-dimensional Bernstein polynomial density estimators. *Metrika*, **41**(3-4), 233–253. MR1293514.

Tenbusch, A. 1997. Nonparametric curve estimation with Bernstein estimates. *Metrika*, **45**(1), 1–30. MR1437794.

Turnbull, B. C., & Ghosh, S. K. 2014. Unimodal density estimation using Bernstein polynomials. *Comput. Statist. Data Anal.*, **72**, 13–29. MR3139345.

Vitale, R. A. 1975. Bernstein polynomial approach to density function estimation. *Pages 87–99 of: Statistical Inference and Related Topics*. Academic Press, New York. MR0397977.